Technical Report 10-001

Modeling of Diffusion through a Network: A New Approach using Cellular Automata and Network Science Techniques

2LT Steven Kinney, Elisha Peterson, Ph.D.

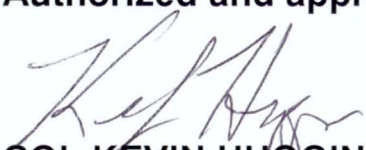
U.S. Military Academy, West Point NY

May 2010

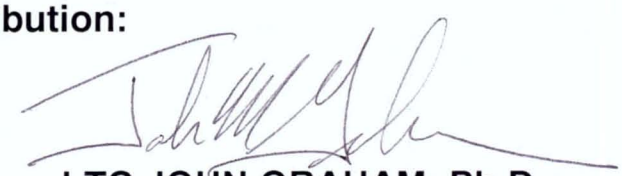
United States Military Academy
Network Science Center

U.S. Military Academy
Network Science Center

Authorized and approved for distribution:



COL KEVIN HUGGINS, Ph.D.
Director of Research



LTC JOHN GRAHAM, Ph.D.
Director

Technical review by

BG(R) Christopher Arney, Ph.D., Department of Mathematical Sciences, U.S. Military Academy

LTC Donovan Phillips, Ph.D., Department of Mathematical Sciences, U.S. Military Academy

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by the U.S. Military Academy Network Science Center. Please address correspondence concerning distribution of reports to: Network Science Center, U.S. Military Academy, 646 Swift Road, West Point, NY 10996

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Military Academy Network Science Center.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
4 May 2010	Technical Report	September 2009 – May 2010

4. TITLE AND SUBTITLE

Modeling of Diffusion through a Network: A New Approach using Cellular Automata and Network Science Techniques

5a. CONTRACT NUMBER
n/a

5b. GRANT NUMBER
n/a

5c. PROGRAM ELEMENT NUMBER
n/a

6. AUTHOR(S)
Steven Kinney, Elisha Peterson

5d. PROJECT NUMBER
ARO NetSci 02

5e. TASK NUMBER
n/a

5f. WORK UNIT NUMBER
n/a

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Network Science Center,
U.S. Military Academy
601 Cullum Road, Thayer Hall Room 119
West Point, NY 10996

8. PERFORMING ORGANIZATION REPORT NUMBER
n/a

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S.Army Research Organization
Triangle Park, NC

10. SPONSOR/MONITOR'S ACRONYM(S)
USMA NSC

11. SPONSOR/MONITOR'S REPORT NUMBER(S) 10-001

12. DISTRIBUTION / AVAILABILITY STATEMENT

Unlimited Distribution

13. SUPPLEMENTARY NOTES
The views expressed in this thesis are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

14. ABSTRACT

This paper investigates the sensitivity of current models that describe diffusion or the spread of infection to changes in topology and/or initial conditions. We construct a computer simulation based on cellular automata that allows for customization of network and infection scenarios. Using the simulation we run several batch runs with varied circumstances to assess the impact of small changes in topology and initial conditions upon diffusion through a network. We also build Excel models to predict long term infection and spread rates, based upon insights from the computer-based cellular automata simulation.

The goal is to highlight weaknesses in current models and present alternative models that may be more realistic. Using network science and computer science techniques, we present another option for modelers to use when forming their own spread and diffusion models.

15. SUBJECT TERMS
Social Network, Graph Theory, Cellular Automata, Disease Modeling, Network Diffusion

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Tish Torgerson
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL	41	**19b. TELEPHONE NUMBER** *(include area code)* 845-938-0804

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Technical Report 10-001

Modeling of Diffusion through a Network: A New Approach using Cellular Automata and Network Science Techniques

2LT Steven Kinney, Elisha Peterson Ph.D.

U.S. Military Academy, West Point NY

U.S. Military Academy Network Science Center
601 Cullum Road, Thayer Hall Room 119, West Point, NY 10996

4 May 2010

ACKNOWLEDGEMENT

The authors are grateful to the Department of Mathematical Sciences at USMA for its support during this project, in particular to Professor Chris Arney and LTC Donovan Phillips for providing valuable feedback on this project.

MODELING OF DIFFUSION THROUGH A NETWORK: A NEW APPROACH USING CELLULAR AUTOMATA AND NETWORK SCIENCE TECHNIQUES

EXECUTIVE SUMMARY

Scientists and mathematicians have been trying for centuries to correctly model behavior of groups of people. Behavior can vary from something as simple as what happens if one member of the group leaves the group to something as complicated as the effect of a natural disaster on a group's cohesion. This paper focuses on the behavior associated with diffusion or spread throughout a group. Exactly what is spreading throughout the group is not defined; it could be a disease or it could be a piece of information. Currently there are several mathematical models that predict how diffusion through a group occurs. Some have been used to predict infection rates in large populations; others have been used to pinpoint individuals that act as key information spreaders such as a local gossiper. The goal of this paper is to offer an alternative method of modeling this diffusion and provide some insight into why this alternative might be more accurate.

Procedure

Many of the current models of diffusion assume random mixing. This would be equivalent to putting a drop of dye in a glass of water and stirring it. The dye will diffuse throughout the entire glass through random movement of the water molecules. Another large segment of these models assume a detailed knowledge of the underlying network connections. For instance an airline company might know exactly which cities had flights connecting them and so would have a detailed understanding of how people diffused around the globe by air. Neither of these two assumptions, random mixing and knowledge of network connections, is always valid. Consider the example of the spread of a disease through a city. To assume that any two individuals in the city have equal chance of spreading the disease as any other two individuals would be an incorrect assumption. Certainly people that go to the same office, the same grocery store, or even live in the same house are more likely to spread disease than two people who do not share any such spaces. To assume detailed knowledge of the network of interactions would also likely be an incorrect assumption. Knowing who interacts with whom is possible in a small office or school class room, but for an entire city the possible number of connections grows too large.

Findings

The tool presented in this paper uses a cellular automata (CA) based model to avoid both invalid assumptions. Instead of assuming random mixing, the CA model assumes random connections. Since connections don't change they better represent relationships and interactions that exist in reality. Also, only the general connectedness of a network needs to be known to apply random connections. Knowing how many interactions people have on average is a lot easier to find than knowing all interactions. Thus the CA model avoids both assumptions, creating a better model in the process.

We validate the CA tool by comparing its output against well known SIS and SIR models. We then go on to use the CA tool to show the important effects that may be masked by assuming random mixing. There are definite variations that occur as a result of the underlying

connections in any network. This tool takes those into effect by keeping the network connections static throughout one simulation period and allowing batch runs so that multiple variations of network connections may be tested under similar conditions and then the results averaged. In the end, this research project successfully concludes with the creation of new method to modeling diffusion through a network and justifying its existence with a critique on existing models.

Academic Presentations without Conference Proceeding

Kinney, S. "Exploration of Diffusion and Infection Models: Sensitivity to Topology and Initial Conditions." SDUMC, St. Francis College, New York, NY (2010).

MODELING OF DIFFUSION THROUGH A NETWORK: A NEW APPROACH USING CELLULAR AUTOMATA AND NETWORK SCIENCE TECHNIQUES

Contents

Equations

Figures

MODELING OF DIFFUSION THROUGH A NETWORK: A NEW APPROACH USING CELLULAR AUTOMATA AND NETWORK SCIENCE TECHNIQUES

I. INTRODUCTION

Network science is an emerging field of study that uses the scientific method to examine an array of networks to derive a series of principles or theorems to describe the behavior of those networks. The U.S. Army recently became interested in network science when the idea of network-centric warfare came about in the late 1990s with the publishing of the book titled the same (Alberts, Garstka, & Stein, 1999). Leading the way in the creation of network-centric operations, the U.S. Army has created a Network Science Center at our very own U.S. Military Academy. Traditionally the focus of network-based operations has been on friendly communication networks and information distribution on the battlefield. This idea stems from the theory that information superiority, just like other more familiar terms (air superiority, fire superiority), is the next step the U.S. Military must work to accomplish in order to dominate in the information age.

Through network science the U.S. Army hopes to develop methods to increase the accuracy, timeliness, and relevance of information that it uses to win the nation's wars. However, network science can also be used to better understand enemy forces. For instance, knowing how the communication networks of the enemy work allows one to target only the most critical components of that network to bring down the whole system. Also, knowing how the enemy communicates provides a tool that can be used to predict the enemy response. So not only does the study of networks afford the U.S. Army greater information sharing abilities, it could also give a better understanding of enemy operations and communications systems.

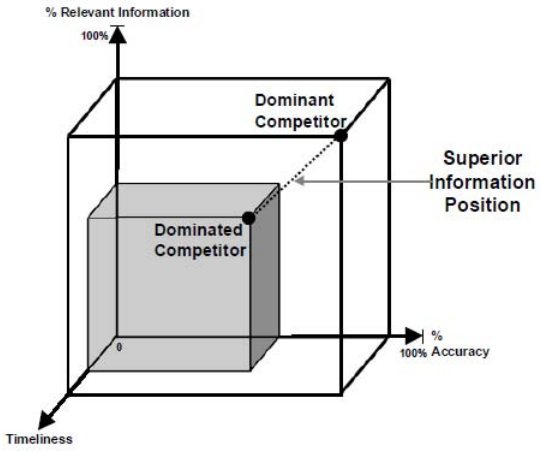


Figure 1 Superior Information Position (From Alberts, Garstka, & Stein, 1999)

Beyond the specific military applications of networks, there are several other possible uses. Like the spread of information, modeling the spread of disease through a population is uniquely suited to study by network science. Modeling the flow of products throughout a consumer network or the delegation throughout a corporate network are also possible areas where network science can offer help. In the information age there are now so many different networks that network study can be applied to all kinds of fields. So while still in its infancy, network science holds the promise of finding unique solutions to complex problems in an increasingly connected world.

Scientists and mathematicians have been trying for centuries to correctly model behavior of groups of people. Behavior can vary from something as simple as what happens if one member of the group leaves the group to something as complicated as the effect of a natural disaster on a group's cohesion. This paper focuses on the behavior associated with diffusion or spread throughout a group. Exactly what is spreading throughout the group is not defined; it could be a disease or it could be a piece of information. Currently there are several mathematical

models that predict how diffusion through a group occurs. Some have been used to predict infection rates in large populations; others have been used to pinpoint individuals that act as key information spreaders such as a local gossiper. The goal of this paper is to offer an alternative method of modeling this diffusion and provide some insight into why this alternative might be more accurate.

Many of the current models of diffusion assume random mixing. This would be equivalent to putting a drop of dye in a glass of water and stirring it. The dye will diffuse throughout the entire glass through random movement of the water molecules. Another large segment of these models assume a detailed knowledge of the underlying network connections. For instance an airline company might know exactly which cities had flights connecting them and so would have a detailed understanding of how people diffused around the globe by air. Neither of these two assumptions, random mixing and knowledge of network connections, is always valid. Consider the example of the spread of a disease through a city. To assume that any two individuals in the city have equal chance of spreading the disease as any other two individuals would be an incorrect assumption. Certainly people that go to the same office, the same grocery store, or even live in the same house are more likely to spread disease than two people who do not share any such spaces. To assume detailed knowledge of the network of interactions would also likely be an incorrect assumption. Knowing who interacts with whom is possible in a small office or school class room, but for an entire city the possible number of connections grows too large.

The tool presented in this paper uses a cellular automata (CA) based model to avoid both invalid assumptions. Instead of assuming random mixing, the CA model assumes random connections. Since connections don't change they better represent relationships and interactions

that exist in reality. Also only the general connectedness of a network needs to be known to apply random connections. Knowing how many interactions people have on average is a lot easier to find than knowing all interactions. Thus the CA model avoids both assumptions, creating a better model in the process.

We validate the CA tool by comparing its output against well known SIS and SIR models. We then go on to use the CA tool to show the important effects that may be masked by assuming random mixing. There are definite variations that occur as a result of the underlying connections in any network. This tool takes those into effect by keeping the network connections static throughout one simulation period and allowing batch runs so that multiple variations of network connections may be tested under similar conditions and then the results averaged. In the end, this research project successfully concludes with the creation of new method to modeling diffusion through a network and justifying its existence with a critique on existing models.

II. BACKGROUND

A. Graph theory

Mathematics has provided a number of different models to describe diffusion through a network. Since networks are generally constructed either to model or facilitate diffusion, being able to describe this diffusion in a numerical expression is of great importance. We begin with a review of some basic ideas in graph theory.

A *graph* is a collection of *vertices* (or *nodes*) and *edges* that connect pairs of vertices known as *neighbors*. A graph may be *directed,* meaning flow from one vertex to another is restricted to only one direction, or it may be *undirected*, in which case flow is unrestricted in direction between nodes. Nodes have a number of attributes that conveniently describe their place within a graph and relative number of neighbors. The *degree* of a node is equal to the number of neighbors that node has. Nodes also typically have some kind of *state* that depends on the type of network being modeled. A disease-tracking graph may have nodes with states such as: infected, immune, not infected, recovered from infection. The undirected graph in Figure 2 has four nodes, with node 1 having a degree of three while node 2 has a degree of one.

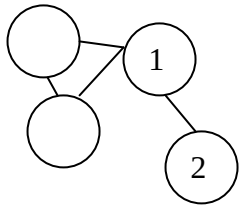


Figure 2 Graph

B. Existing Diffusion Models

1. Threshold v. Independent Cascade Models[1]

Two general types of diffusion models are the *threshold* and *independent cascade* models. These models are named for the conditions by which a node changes state. The first model gives each node its own threshold which must be reached before it will change state. The independent cascade model gives each node a single chance to change the state of neighboring nodes; each node is acting independently of the others.

In the threshold model each neighbor may be given a certain weight with which it will affect its neighbor(s). For a particular node to change state the combined weight of that node's neighbors acting on it must meet the threshold, as specified in Inequality 1.

$$\sum_{degree\ of\ w} b_{v,w} \leq \theta_w$$

Inequality 1 Threshold Weights (From Kempe, Kleinberg, & Tardos, 2003)

In this summation, the weights (*b*) of the neighbors (*v*) of *w* are added together. If they are less than or equal to the threshold θ of the node *w* then the node will not change state. This computation is done for each node at each time step (*t)*.

The second type of model, the independent cascade model, simply gives each of the neighbors of *v* some independent probability of affecting the state of *v*. For example if a neighbor of *w* called *v* becomes active at time *t*, it is given a single chance to activate *w* with a random probability *p(v)*. If *v* succeeds, then *w* becomes active at time *t* + 1. If there are multiple

[1] Kempe, Kleinberg, & Tardos, 2003, 138

neighbors of *w* that become active at time *t* then they are all given a chance to activate *w* in a random order.

2. Bass Model[2]

One of the earlier models of diffusion is known as the Bass Model (Jackson, 2008). This model does not involve network science but is a commonly used and well known model of diffusion. The model depends on two parameters: the first is the rate (*p*) at which nodes spontaneously become active and the second is the rate (*q*) at which nodes become active due to the activity of neighboring nodes. For simplicity it can be assumed that there are only two states, active and inactive. The model predicts *F(t)*, the number of active nodes at time t by the following equation:

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}}$$

Equation 2 Bass Model (From Jackson, 2008)

In this equation, *p* is the rate of spontaneous activity and *q* is the rate of activation due to surrounding activity. Using this model the process of calculating an approximation for the total number of active nodes at any time is quite straight forward.

3. SIS & SIR Model[3]

Another diffusion model known as the SIS model (Susceptible, Infected, and Susceptible) is more commonly associated with modeling the spread of disease. This model is based on a

[2] Jackson, 2008, 187

[3] Ibid., 196.

node being either infected or not infected but susceptible to further infection. This model differs from the SIR model (Susceptible, Infected, and Recovered) in that the SIR model incorporates a state in which a node is no longer susceptible to infection either through immunity or death. Important to both of these models is the fact that random mixing of individuals is assumed, so that there are no predefined pathways for disease spread. The SIS model is the simpler of the two since all nodes return to their original state after infection. Its measure of the average infection rate is rather simple to calculate and depends on two factors. The first is the degree distribution *P*, and the second is the fraction of individuals of degree *d* who are infected, *ρ(d)*. Then, the average infection rate of the population is

$$\rho = \sum_{d=0,1,2\ldots} P(d)\rho(d).$$

Equation 3 SIS Model (From Jackson, 2008)

The SIR model, of which the Kermack-McKendrick is a common type, is a little more complex. It consists of a system of nonlinear ordinary differential equations, one for each state a node could be in: S (susceptible), I (infected/infectious), and R (removed)

$$\frac{dS(t)}{dt} = bS(t) - vI(t)S(t)$$

$$\frac{dI(t)}{dt} = vI(t)S(t) - cI(t)$$

$$\frac{dR(t)}{dt} = cI(t)$$

Equation 4 SIR Equations (From Weisstein)

where v is the *infection* rate, b is the *birth* rate, and c is the *immunity* rate. The behavior of this model depends upon the ratio of the initially infected (S) times the infection rate (β) to the recovery rate (γ),

$$R_0 = \frac{\beta S}{\gamma}.$$

Equation 5 SIR Ratio (From Weisstein)

If this ratio is greater than 1, then each person who is infected will infect more than one other person and thus the disease will spread. If it is less than 1, then the disease will die out quickly.

C. Cellular Automata

While the concept of cellular automata has no immediate relationship with diffusion through a network it does have some specific applications to our particular objectives. A *cellular automata* (CA) consists of a collection of cells that change over discrete time intervals according to a system of rules. These rules are applied at each time step to decide the state at the next time step. Each time step is known as a generation. Often changes in the cells are depicted in the form of coloration, in its simplest form this may be a change from black to white or vice-a-versa. Cells may be in a single line representing a one-dimensional environment, or they may be placed in a grid to simulate a two-dimensional situation. There are no universal generic rules for cellular automata, but often particular rules are applied in a manner similar to threshold models described in the section 1. The state of a cell in a future generation often depends upon the cell's current state and the count of states of neighboring cells. In CA, neighboring cells are generally defined as those cells directly adjacent to the cell in question. For instance in a two state system a cell may enter state 2 in the next generation if in the previous generation the cell was in state 1 and had two neighbors in state 2.

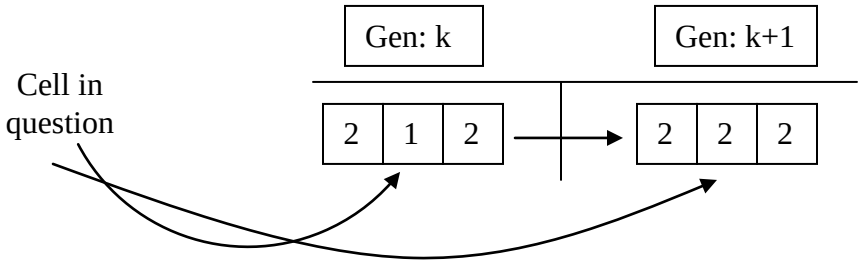


Figure 3 Cellular Automata

D. Objective and Research Question

The objective of this paper is to describe the design and implementation of a new method for modeling spread of information. This method will seek to combine elements of both network graphs and geographically-based models, building upon the theory of cellular automata. CA is useful because the underlying network structure that connects cells and the time step logic enables easy modeling of large or complex situations.

We use a custom-made form of CA, and interlace some ideas from network graphs such as "wormhole" links that connect cells that are not adjacent. We also introduce some degree of randomness into the system using probability distributions to determine whether information will flow from a cell to its neighbor.

We verify the validity of the model by comparing it to current models already mentioned in the above before using it to draw other conclusions about the diffusion in a network. The final objective is the formulation of conclusions about diffusion through networks, which can be used to evaluate current models with data generated by the CA tool.

E. Justification of Study

In the Army, or indeed in any organization, communication is an aspect of leadership that is highly valued. Much research has been done and money invested into the study of information flow and communication between individuals and within a large group. Especially in the realm of network science, the study of entities interacting as part of a larger group has unearthed several useful models or methods for information spread. While these models often produce useful and interesting results, they can be difficult to construct based on incomplete knowledge of the edges in the network graph. Strictly geography-based models used for modeling spread simply by proximity of the entities to one another fail to account for long-distance communications that allow some entities to share information at a distance.

Outside of network science there are many models that don't take into account any underlying network when modeling diffusion. It is one of the goals of this paper to underscore the importance of recognizing the underlying graph of a network or at the very least taking into account some basic properties of that graph. One of the more visually appealing as well as easily implemented methods for modeling spread geographically is through the use of CA. It is for these reasons that a combination of CA and network theory will create a unique and useful model of diffusion.

III. MODELING TOOLS

A major goal of this research project and one that took a vast amount of the time available to complete it was the creation of a modeling tool from which to gather data. The initial idea was to create a CA-based computer simulator that was customizable and easy to use. Customization was easy to accomplish since the tool was built from the ground up and the ease of use came mostly out of the GUI interface. Once the CA tool had been created and modified to an acceptable point of operation, we created another Excel-based tool. This experimental tool was based off of the recently constructed CA tool but instead of discrete states of 0 and 1 in CA, its computations were based on probabilities.

A. Cellular Automata Tool

The CA tool, pictured in Figure 4, is designed to be easy to use and highly customizable. It allows the user to modify the size of the CA universe and the number of generations to run. It can run batch runs of various scenarios to allow easy computation of averages. The CA rules can be modified to model different scenarios such as SIS or SIR scenarios. Even the connections between cells can be modified using the available adjacency matrix or the default settings of left and right adjacent neighbor connections may be used.

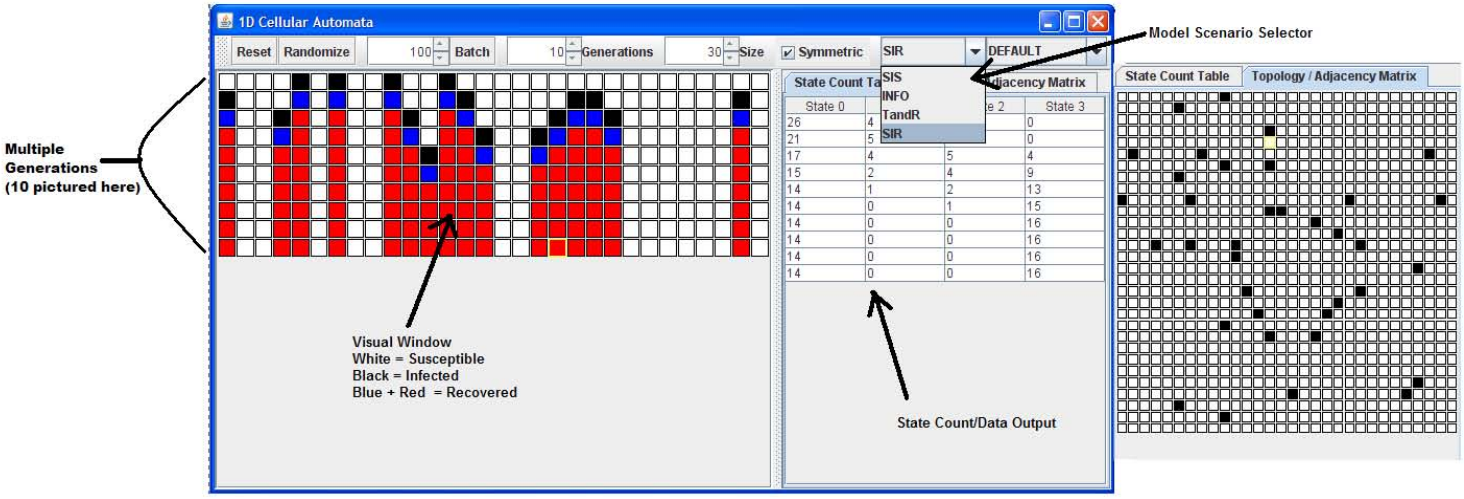
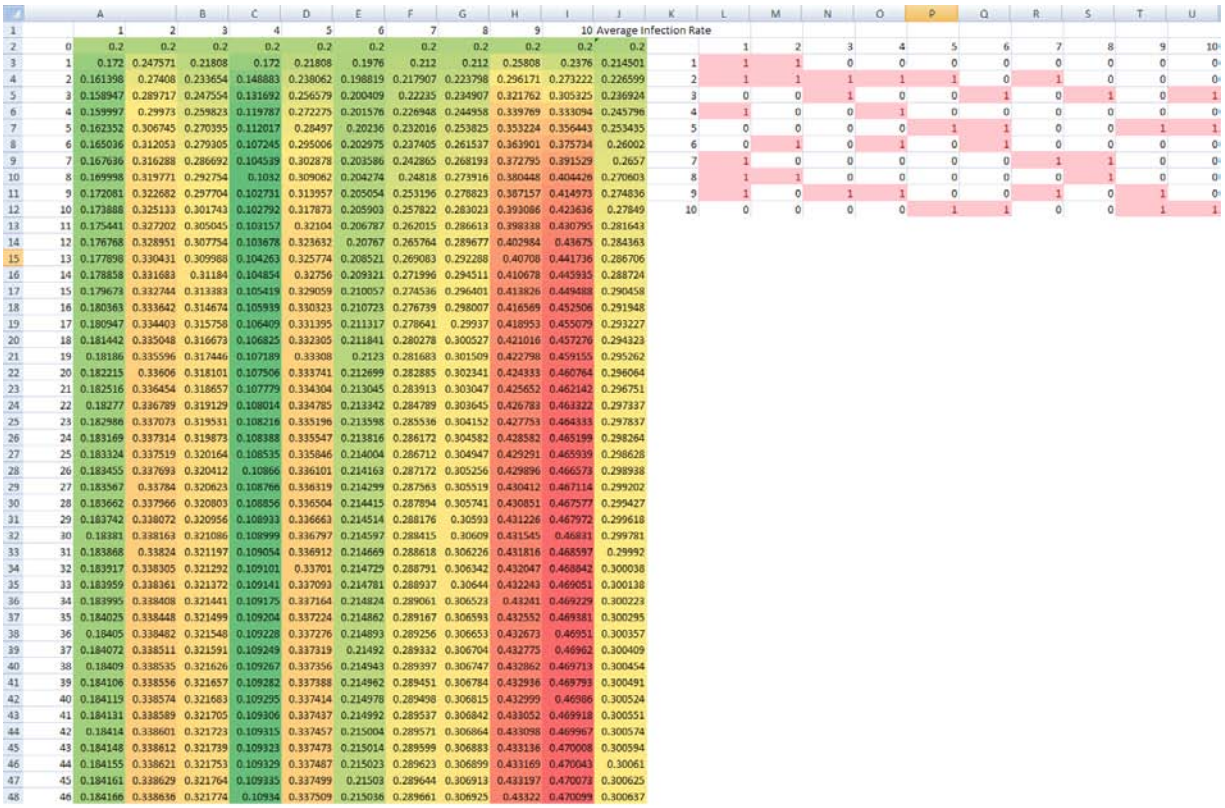


Figure 4 Cellular Automata Tool

In the visual window each successive generation is displayed in the row below. Likewise the state count window displays the total number of cells in each state in each generation by row. In the adjacency table, filled in squares represent connections between cells in the visual window and checking/unchecking the symmetric box can enforce symmetry in the adjacency matrix thus making either a directed or undirected graph. Depending on the scenario, the adjacency table may be either in a default configuration with left and right neighbor connections only, or may be populated with a random fill of connections according to the level of connectedness desired. Also, the default configuration may have additional random connections according to the desired level of connectedness. The entire tool was built from scratch within Java, allowing for complete customization.

B. Experimental Excel Tool

The Excel model was created to explore specific questions inspired by the CA tool. Excel is especially well-suited for CA modeling because it can easily be used to perform repeated operations over a finite and discrete amount of time. This model focused on probabilities of infection rather than discrete cell states of infected or uninfected, in order to examine the approximate long-term behavior of the system. Figure 5 below depicts an example simulation of a population of ten individuals. The initial infection rate was 20% and the chance of transmission and recovery were 20% and 30% respectively. The adjacency matrix in the upper right corner of the figure operates similarly to the matrix in the CA tool. The colors of the columns indicate their infection rate with red being more likely to become infected and green less likely.



	A	B	C	D	E	F	G	H	I	J
1	1	2	3	4	5	6	7	8	9	10 Average Infection Rate
2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
3	0.172	0.247571	0.21808	0.172	0.21808	0.1976	0.212	0.212	0.25808	0.2376
4	0.161398	0.27408	0.233654	0.148883	0.238062	0.198819	0.217907	0.223798	0.296171	0.273222
5	0.158947	0.289717	0.247554	0.131692	0.256579	0.200409	0.22235	0.234907	0.321762	0.305325
6	0.159997	0.29973	0.259823	0.119787	0.272275	0.201576	0.226948	0.244958	0.339769	0.333094
7	0.162352	0.306745	0.270395	0.112017	0.28497	0.20236	0.232016	0.253825	0.353224	0.356443
8	0.165036	0.312053	0.279305	0.107245	0.295006	0.202975	0.237405	0.261537	0.363901	0.375734
9	0.167636	0.316288	0.286692	0.104539	0.302878	0.203586	0.242865	0.268193	0.372795	0.391529
10	0.169998	0.319771	0.292754	0.1032	0.309062	0.204274	0.24818	0.273916	0.380448	0.404426
11	0.172081	0.322682	0.297704	0.102731	0.313957	0.205054	0.253196	0.278823	0.387157	0.414973
12	0.173888	0.325133	0.301743	0.102792	0.317873	0.205903	0.257822	0.283023	0.393086	0.423636
13	0.175441	0.327202	0.305045	0.103157	0.32104	0.206787	0.262015	0.286613	0.398338	0.430795
14	0.176768	0.328951	0.307754	0.103678	0.323632	0.20767	0.265764	0.289677	0.402984	0.43675
15	0.177898	0.330431	0.309988	0.104263	0.325774	0.208521	0.269083	0.292288	0.40708	0.441736
16	0.178858	0.331683	0.31184	0.104854	0.32756	0.209321	0.271996	0.294511	0.410678	0.445935
17	0.179673	0.332744	0.313383	0.105419	0.329059	0.210057	0.274536	0.296401	0.413826	0.449488
18	0.180363	0.333642	0.314674	0.105939	0.330323	0.210723	0.276739	0.298007	0.416569	0.452506
19	0.180947	0.334403	0.315758	0.106409	0.331395	0.211317	0.278641	0.29937	0.418953	0.455079
20	0.181442	0.335048	0.316673	0.106825	0.332305	0.211841	0.280278	0.300527	0.421016	0.457276
21	0.18186	0.335596	0.317446	0.107189	0.33308	0.2123	0.281683	0.301509	0.422798	0.459155
22	0.182215	0.33606	0.318101	0.107506	0.333741	0.212699	0.282885	0.302341	0.424333	0.460764
23	0.182516	0.336454	0.318657	0.107779	0.334304	0.213045	0.283913	0.303047	0.425652	0.462142
24	0.18277	0.336789	0.319129	0.108014	0.334785	0.213342	0.284789	0.303645	0.426783	0.463322
25	0.182986	0.337073	0.319531	0.108216	0.335196	0.213598	0.285536	0.304152	0.427753	0.464333
26	0.183169	0.337314	0.319873	0.108388	0.335547	0.213816	0.286172	0.304582	0.428582	0.465199
27	0.183324	0.337519	0.320164	0.108535	0.335846	0.214004	0.286712	0.304947	0.429291	0.465939
28	0.183455	0.337693	0.320412	0.10866	0.336101	0.214163	0.287172	0.305256	0.429896	0.466573
29	0.183567	0.33784	0.320623	0.108766	0.336319	0.214299	0.287563	0.305519	0.430412	0.467114
30	0.183662	0.337966	0.320803	0.108856	0.336504	0.214415	0.287894	0.305741	0.430851	0.467577
31	0.183742	0.338072	0.320956	0.108933	0.336663	0.214514	0.288176	0.30593	0.431226	0.467972
32	0.18381	0.338163	0.321086	0.108999	0.336797	0.214597	0.288415	0.30609	0.431545	0.46831
33	0.183868	0.33824	0.321197	0.109054	0.336912	0.214669	0.288618	0.306226	0.431816	0.468597
34	0.183917	0.338305	0.321292	0.109101	0.33701	0.214729	0.288791	0.306342	0.432047	0.468842
35	0.183959	0.338361	0.321372	0.109141	0.337093	0.214781	0.288937	0.30644	0.432243	0.469051
36	0.183995	0.338408	0.321441	0.109175	0.337164	0.214824	0.289061	0.306523	0.43241	0.469229
37	0.184025	0.338448	0.321499	0.109204	0.337224	0.214862	0.289167	0.306593	0.432552	0.469381
38	0.18405	0.338482	0.321548	0.109228	0.337276	0.214893	0.289256	0.306653	0.432673	0.46951
39	0.184072	0.338511	0.321591	0.109249	0.337319	0.21492	0.289332	0.306704	0.432775	0.46962
40	0.18409	0.338535	0.321626	0.109267	0.337356	0.214943	0.289397	0.306747	0.432862	0.469713
41	0.184106	0.338556	0.321657	0.109282	0.337388	0.214962	0.289451	0.306784	0.432936	0.469793
42	0.184119	0.338574	0.321683	0.109295	0.337414	0.214978	0.289498	0.306815	0.432999	0.46986
43	0.184131	0.338589	0.321705	0.109306	0.337437	0.214992	0.289537	0.306842	0.433052	0.469918
44	0.18414	0.338601	0.321723	0.109315	0.337457	0.215004	0.289571	0.306864	0.433098	0.469967
45	0.184148	0.338612	0.321739	0.109323	0.337473	0.215014	0.289599	0.306883	0.433136	0.470008
46	0.184155	0.338621	0.321753	0.109329	0.337487	0.215023	0.289623	0.306899	0.433169	0.470043
47	0.184161	0.338629	0.321764	0.109335	0.337499	0.21503	0.289644	0.306913	0.433197	0.470073
48	0.184166	0.338636	0.321774	0.10934	0.337509	0.215036	0.289661	0.306925	0.43322	0.470099

	K	L	M	N	O	P	Q	R	S	T	U
		1	2	3	4	5	6	7	8	9	10
1		1	1	0	0	0	0	0	0	0	0
2		1	1	1	1	1	0	1	0	0	0
3		0	0	1	0	0	1	0	1	0	1
4		1	0	0	1	0	0	0	0	0	0
5		0	0	0	0	1	1	0	0	1	1
6		0	1	0	1	0	1	0	0	0	0
7		1	0	0	0	0	0	1	1	0	0
8		1	1	0	0	0	0	0	1	0	0
9		1	0	1	1	0	0	1	0	1	0
10		0	0	0	0	1	1	0	0	1	1

Figure 5 Excel Model of Infection

The entire model was built using formulas so that the initial infection rate, recovery and transmission rates, as well as the percent connectedness of the network could easily be changed to affect the whole model. There is no batch run ability for this model so computing overall averages for a given scenario is not possible. There were some interesting results, however, which will be examined in the findings section of this paper.

IV. Results

A. Validation

We begin our discussion of results with a validation of the tool being used to model diffusion. We will use the SIS and SIR models described above and compare the output to data from computer simulations using our CA tool. This simplest of the two models (SIS) can generally be described by the figure below:

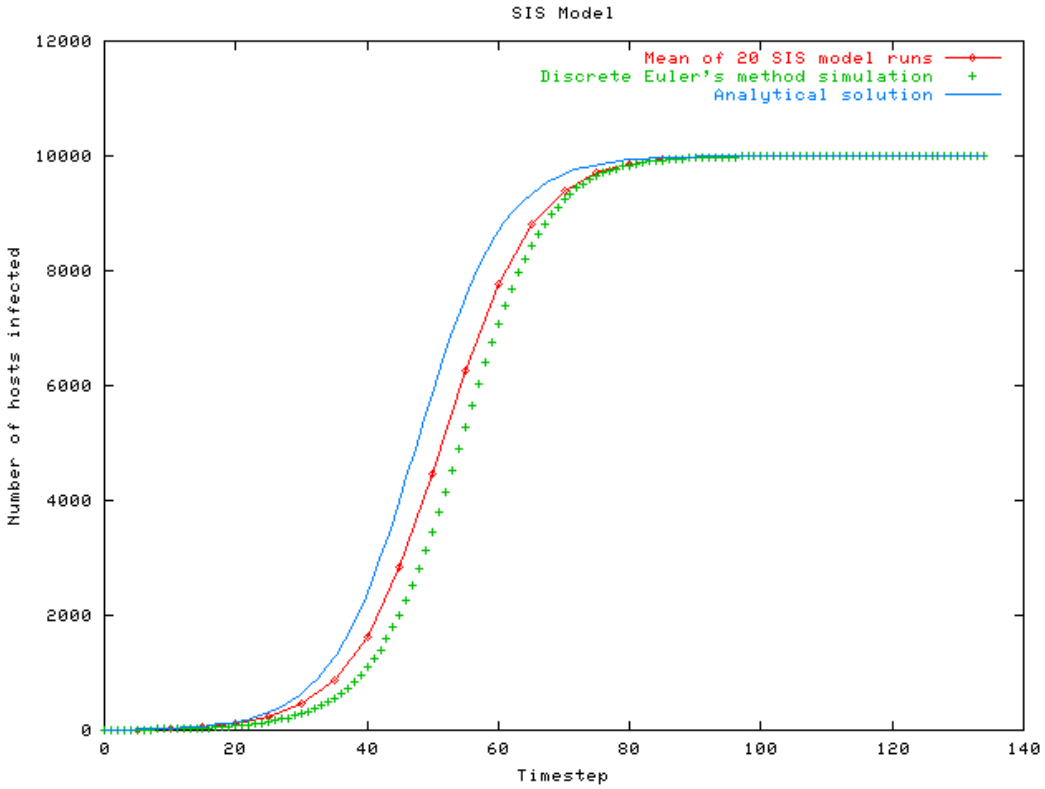


Figure 6 Simulation and Solutions to SIS Model (From Ediger, 2010)

This figure depicts a simulation average, a numerical solution, and an analytical solution to the SIS model measuring the number of infected nodes over time. As you can see, the curve exhibits a logistical pattern. Similar results were obtained from an SIS model simulation using our computer based CA:

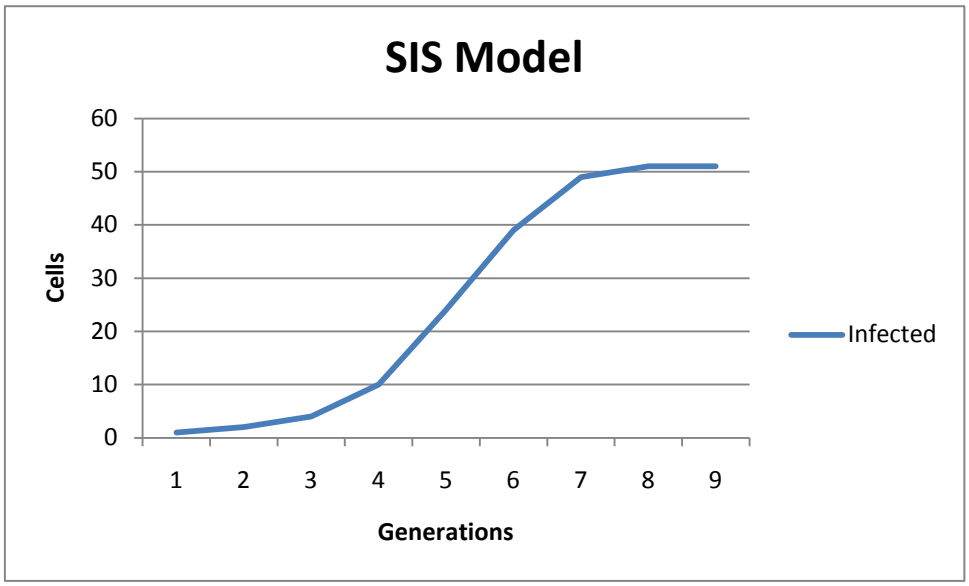


Figure 7 CA Simulation of SIS System

This simulation was created using two state cells (0 & 1) with neighboring cell connections and additional random connections on the order of 0.5% of the total possible connections and a universe of 100 cells. The simulation was run for 50 generations but is cut off in this graph as there was no change in cell count after 8 generations. SIS rules were in place that allowed an infected cell to transfer its infection to any connected cells and then become susceptible in subsequent time steps.

Both the well known SIS model and the simulation run using our CA tool give similar logistical patterns of infection growth rates. These curves correctly model an SIS sytem as the infection starts off in exponential growth initially and then slowly levels off as the infection rate nears the carrying capacity of the disease. The similarity of these two curves indicates that our CA tool is a valid method of modeling diseases in an SIS system.

Next, the tool was validated by modeling a more complicated model, the SIR model. Curves for a SIR system as described in the previous section generally appear similar to the curve below:

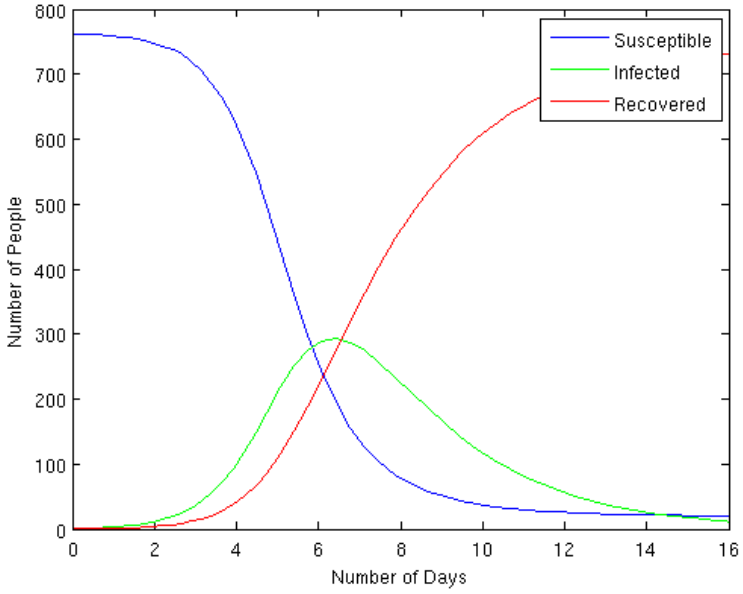


Figure 8 Deterministic Solution of SIR Model (From Chestnut, 2010)

This image was generated from a deterministic solution to the three differential equations that make up the SIR model using MATLAB, starting with 1 infected individual and over 700 susceptible individuals. Below are the results of an SIR simulation using the CA tool:

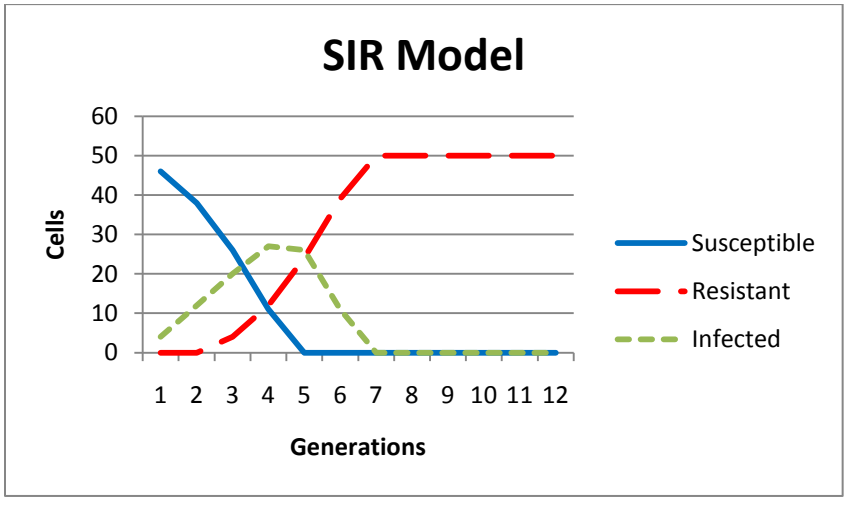


Figure 9 CA Simulation of SIR System

This simulation was created using three states of cells with neighboring cell connections and a universe of 50 cells. SIR rules were in place that allowed a susceptible cell to become infected by any neighboring cell, remaining infected for a generation, and then transforming into a resistant cell for the remainder of the time period.

The deterministic solution and the CA simulation are similar in appearance. Both figures correctly model the initial rapid increase in infections resulting decrease in susceptibility and follow on increase in resistance to infection. The infection rate slows and drops off as the number of resistant individuals/cells becomes greater than the amount of susceptible ones. Both the infection rate and susceptible rate approach zero as all individuals/cells become resistant to the infection.

One important aspect to note about these two models is that they both are based on the assumption of random mixing. CA does not allow for random mixing to occur. However, the CA universe is set at the beginning of a run and does not change throughout the period of a simulation. This is an important distinction to make as random mixing cannot always be assumed. In reality it cannot be expected that an individual will spread an infection to an area to which that individual has never been to. Disease is spread along channels of contacts and so the spread of a disease is best modeled using an underlying graph of the connections between people.

B. Findings

Spread of Information through a Community

Obviously it is difficult if not impossible to know the underlying graph of a large network especially if the nodes of that network are as uncontrollable as people can be. It is unrealistic to expect a modeler to know all of the connections of such a graph but one could know the average number of connections between nodes. The CA tool used in this paper allows for the customization of the connectedness of the graph by randomly apportioning a percentage of the total possible connections. This is not the same as random mixing: once the simulation begins the connections do not change. Depending on the real life situation being modeled this may or may not be a more accurate picture of what is being modeled.

One of the situations the CA tool was used to model that showed the importance of the topology, was in the modeling of information flow from person to person through a community. Using a few assumptions to create a very simple scenario, we modeled the flow of information with a 100% chance of transmittance between 100 connected individuals. We then varied the number of connections and the distribution of those connections to examine how the topology of a network affected the diffusion.

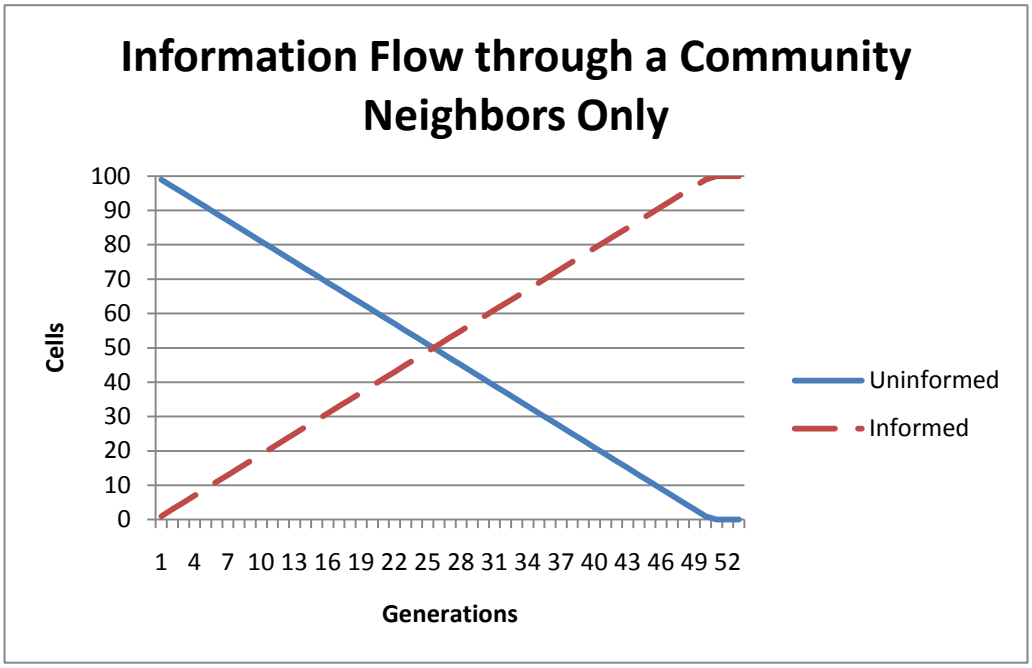


Figure 10 Information Flow through a Community Neighbors Only

The very uninteresting figure above shows the flow of information from one individual through an entire community where each individual may transfer information to their two immediately adjacent neighboring individuals. This would best be envisioned as a circle of 100 family houses where each family only communicated to the families directly to the right and left of their own house. This is not a very realistic model but it provides a base line to compare things to.

The next simulation eliminated the rigid left and right neighborhood structure and replaced it with a random assortment of connections that consisted of 2% of the total possible number of connections amongst the individual families. Since there are 100 families there are 10,000 possible connections, thus 2% of the possible connections represent 200 connections which is the same number of connections in the simulation just prior with left and right neighbors. In this simulation the number of connections is the same as the one before but the

placement of those connections is not. The average result of this model of information flow through a community (for 500 runs) is depicted in Figure 9 below.

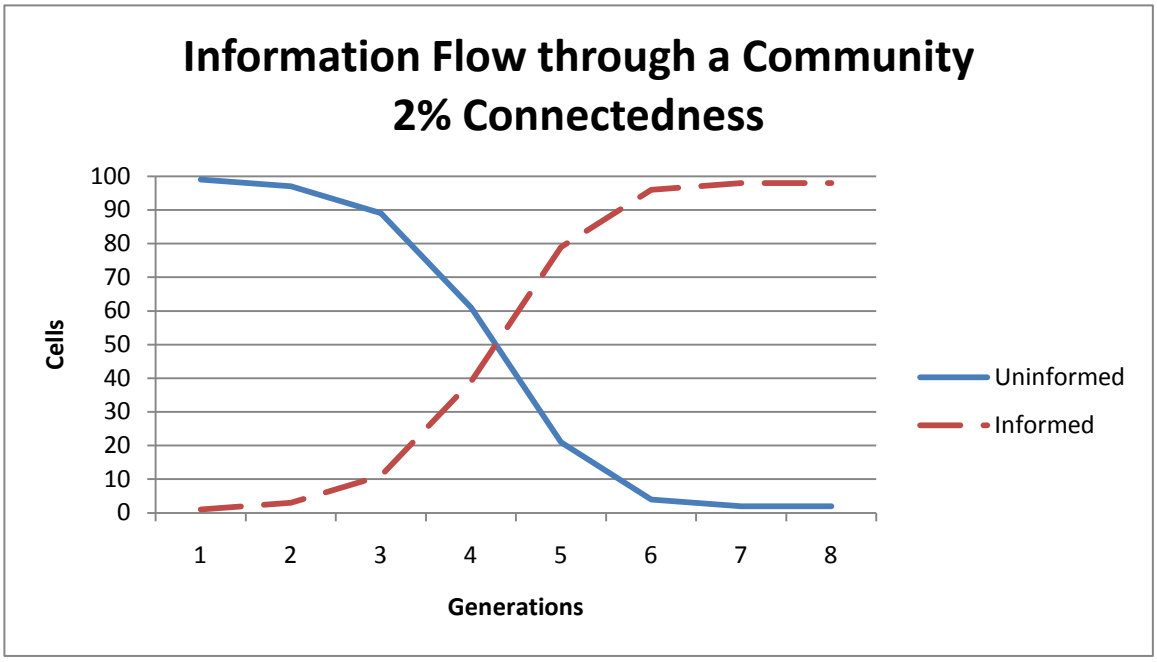


Figure 11 Information Flow through a Community 2% Connectedness

In this figure one does not see the same linear growth that occurred in the scenario depicted by Figure 8. Additionally notice that the number of informed families does not nor will it reach 100% of the total. This means that the entire community does not always receive the information that is being passed along. Sometimes there are families that are isolated from the rest of the community and thus cannot receive the information from anyone else in the community. Also the time scale is much shorter in this scenario. Whereas in the previous scenario it took 50-51 generations for the entire community to become informed, in this scenario it took only eight generations for almost all of the community to become informed. Given these results, it appears that randomizing the connections between families increased the rate of transference but reduced the possibility of the entire community becoming informed.

Reducing the number of connections to 1% of the total possible connections significantly reduced the chance of the entire community from becoming informed but did not increase the rate at which transference occurred. In Figure 12 below the results of an averaged 500 batch runs of a community randomly connected with 1% of the total possible connections are depicted.

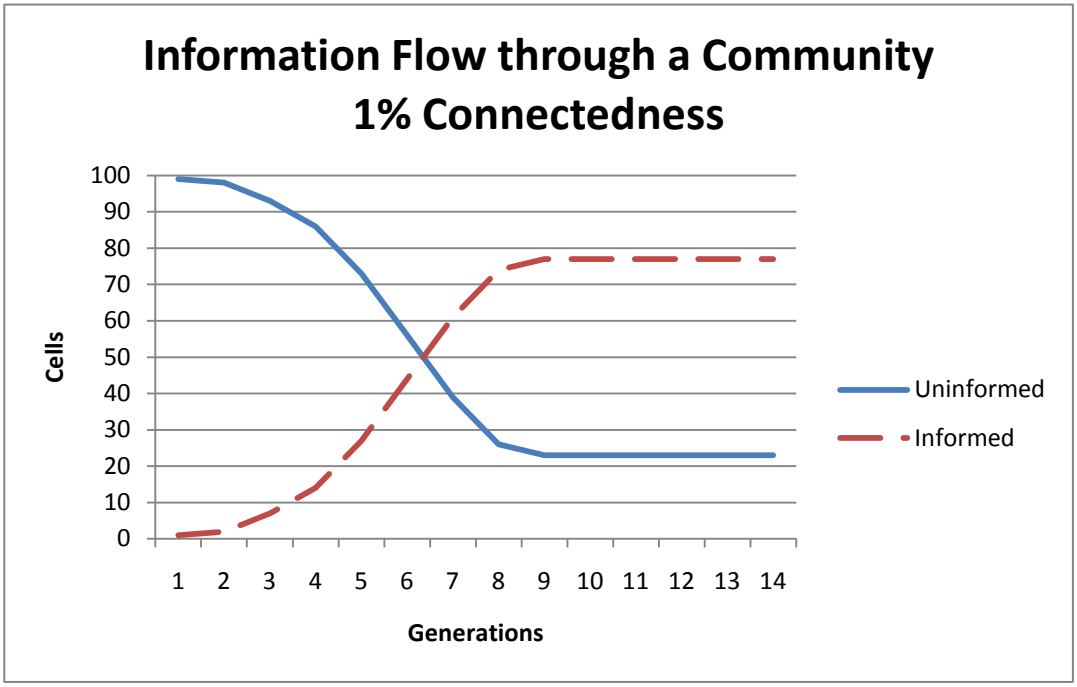


Figure 12 Information Flow through a Community 1% Connectedness

The smaller number of connections increased the likelihood that some families would be entirely disconnected from the group of the community that was connected to the one informed family. The rate at which the community reached maximum information saturation occurred after 9 generations, which was not really different from the rate at which the 2% connectedness simulation reached saturation.

Disease Transmission and Expected Infection Rates

Another scenario tested was a threshold based model that was a variant of the SIS model. Like the SIS model, each cell had two states, infected and susceptible, but there were also transmission and recovery rates associated with each transformation of cell state. We initially choose a transmission rate of 35% and a recovery rate of 50% because they gave significant variability in the results. Figure 13 below depicts simulations using left and right neighbor connections, and additional 0.0%, 0.50%, 1.0% or 1.50% connections determined at random. Each simulation is given its own color with the number of infected individuals starting at 10% and the number susceptible at 90% with a total population of 100 cells.

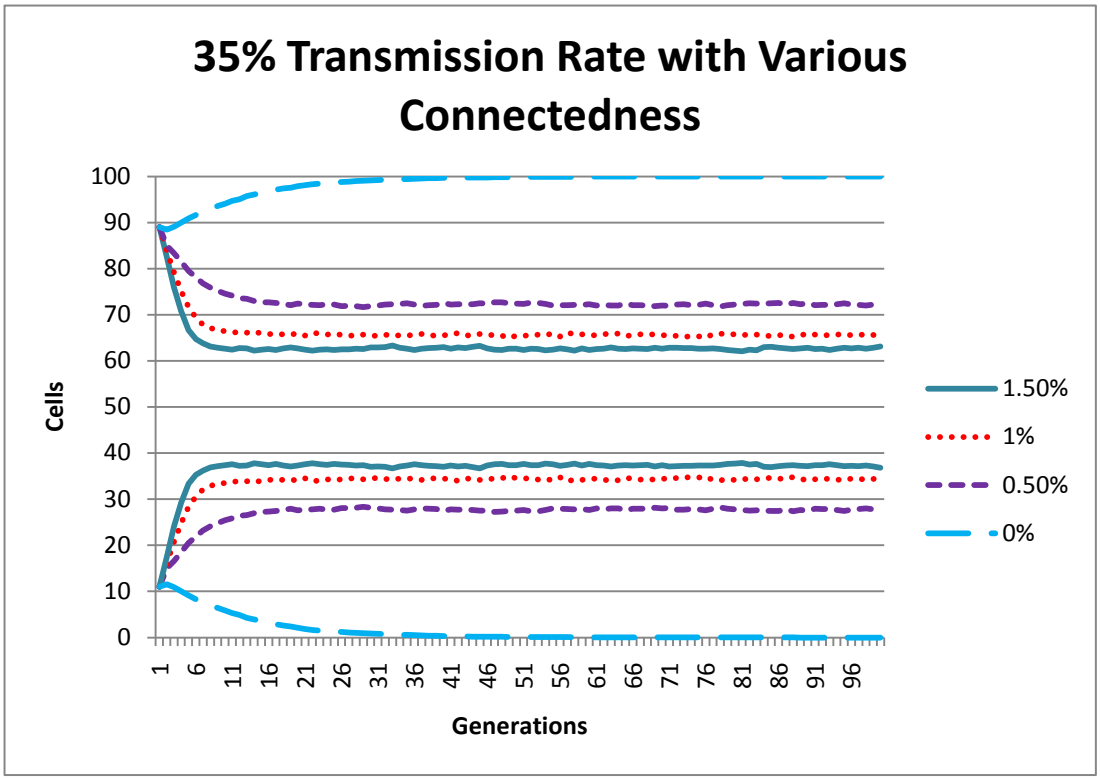


Figure 13 35% Transmission Rate & Various Connectedness

From these results it appears that higher connectedness leads to higher rates of population infection rates. What is surprising by these results is how big of a difference the added

connections made in the infection rates. With just neighbor connections (0.0%) the infection died off fairly quickly since recovery was more likely than transmission. But added connections quickly overcame the difference in transmission and recovery rates allowing the infection to have a long term presence in the population. Traditional models predict that with random mixing a disease will die off if it has a significantly smaller transmission rate compared to recovery rate, see equation 4 (SIR Ratio). But the CA simulation shows that the infection rate is highly dependent upon the connectedness of the network; in fact, a disease may still persist in a population with a transmission rate smaller than a recovery rate.

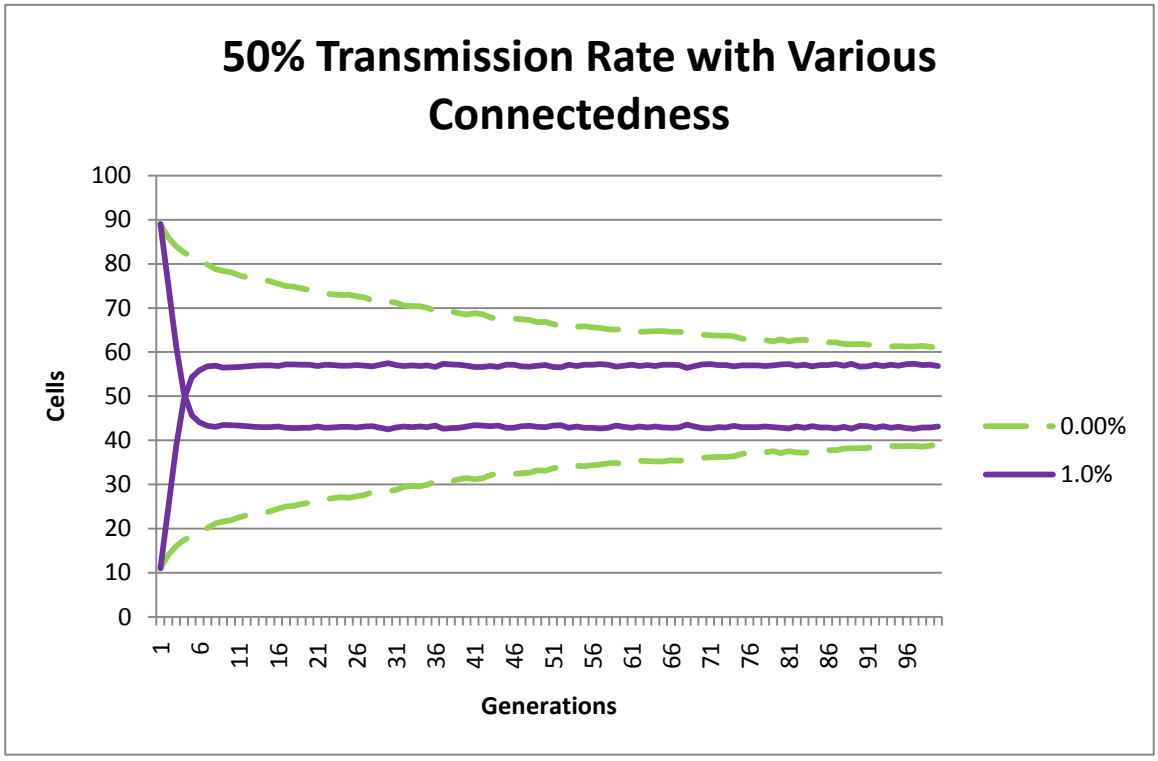


Figure 14 50% Transmission Rate & Various Connectedness

When reversed, transmission rate of 50% and recovery rate of 35%, the expected persistant infection rates emerge even with just left and right neighbor connections. The additional random connections pushes the infection rate higher than 50%. These two scenarios

show the immense effect that the connectedness of a network can have on the spread of a pathogen through a community.

Excel Model Examples

Figure 5 (below again for reference) of the Excel Model of Infection showed clearly that some individuals were more likely to become infected than others, even when all individuals in a population initial have the same likelihood of infection.

	A	B	C	D	E	F	G	H	I	J
1	1	2	3	4	5	6	7	8	9	10 Average Infection Rate
2	0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
3	1	0.172	0.247571	0.21808	0.172	0.21808	0.1976	0.212	0.212	0.25808	0.2376	0.214501
4	2	0.161398	0.27408	0.233654	0.148883	0.238062	0.198819	0.217907	0.223798	0.296171	0.273222	0.226599
5	3	0.158947	0.289717	0.247554	0.131692	0.256579	0.200409	0.22235	0.234907	0.321762	0.305325	0.236924
6	4	0.159997	0.29973	0.259823	0.119787	0.272275	0.201576	0.226948	0.244958	0.339769	0.333094	0.245796
7	5	0.162352	0.306745	0.270395	0.112017	0.28497	0.20236	0.232016	0.253825	0.353224	0.356443	0.253435
8	6	0.165036	0.312053	0.279305	0.107245	0.295006	0.202975	0.237405	0.261537	0.363901	0.375734	0.26002
9	7	0.167636	0.316288	0.286692	0.104539	0.302878	0.203586	0.242865	0.268193	0.372795	0.391529	0.2657
10	8	0.169998	0.319771	0.292754	0.1032	0.309062	0.204274	0.24818	0.273916	0.380448	0.404426	0.270603
11	9	0.172081	0.322682	0.297704	0.102731	0.313957	0.205054	0.253196	0.278823	0.387157	0.414973	0.274836
12	10	0.173888	0.325133	0.301743	0.102792	0.317873	0.205903	0.257822	0.283023	0.393086	0.423636	0.27849
13	11	0.175441	0.327202	0.305045	0.103157	0.32104	0.206787	0.262015	0.286613	0.398338	0.430795	0.281643
14	12	0.176768	0.328951	0.307754	0.103678	0.323632	0.20767	0.265764	0.289677	0.402984	0.43675	0.284363
15	13	0.177898	0.330431	0.309988	0.104263	0.325774	0.208521	0.269083	0.292288	0.40708	0.441736	0.286706
16	14	0.178858	0.331683	0.31184	0.104854	0.32756	0.209321	0.271996	0.294511	0.410678	0.445935	0.288724
17	15	0.179673	0.332744	0.313383	0.105419	0.329059	0.210057	0.274536	0.296401	0.413826	0.449488	0.290458
18	16	0.180363	0.333642	0.314674	0.105939	0.330323	0.210723	0.276739	0.298007	0.416569	0.452506	0.291948
19	17	0.180947	0.334403	0.315758	0.106409	0.331395	0.211317	0.278641	0.29937	0.418953	0.455079	0.293227
20	18	0.181442	0.335048	0.316673	0.106825	0.332305	0.211841	0.280278	0.300527	0.421016	0.457276	0.294323
21	19	0.18186	0.335596	0.317446	0.107189	0.33308	0.2123	0.281683	0.301509	0.422798	0.459155	0.295262
22	20	0.182215	0.33606	0.318101	0.107506	0.333741	0.212699	0.282885	0.302341	0.424333	0.460764	0.296064
23	21	0.182516	0.336454	0.318657	0.107779	0.334304	0.213045	0.283913	0.303047	0.425652	0.462142	0.296751
24	22	0.18277	0.336789	0.319129	0.108014	0.334785	0.213342	0.284789	0.303645	0.426783	0.463322	0.297337
25	23	0.182986	0.337073	0.319531	0.108216	0.335196	0.213598	0.285536	0.304152	0.427753	0.464333	0.297837
26	24	0.183169	0.337314	0.319873	0.108388	0.335547	0.213816	0.286172	0.304582	0.428582	0.465199	0.298264
27	25	0.183324	0.337519	0.320164	0.108535	0.335846	0.214004	0.286712	0.304947	0.429291	0.465939	0.298628
28	26	0.183455	0.337693	0.320412	0.10866	0.336101	0.214163	0.287172	0.305256	0.429896	0.466573	0.298938
29	27	0.183567	0.33784	0.320623	0.108766	0.336319	0.214299	0.287563	0.305519	0.430412	0.467114	0.299202
30	28	0.183662	0.337966	0.320803	0.108856	0.336504	0.214415	0.287894	0.305741	0.430851	0.467577	0.299427
31	29	0.183742	0.338072	0.320956	0.108933	0.336663	0.214514	0.288176	0.30593	0.431226	0.467972	0.299618
32	30	0.18381	0.338163	0.321086	0.108999	0.336797	0.214597	0.288415	0.30609	0.431545	0.46831	0.299781
33	31	0.183868	0.33824	0.321197	0.109054	0.336912	0.214669	0.288618	0.306226	0.431816	0.468597	0.29992
34	32	0.183917	0.338305	0.321292	0.109101	0.33701	0.214729	0.288791	0.306342	0.432047	0.468842	0.300038
35	33	0.183959	0.338361	0.321372	0.109141	0.337093	0.214781	0.288937	0.30644	0.432243	0.469051	0.300138
36	34	0.183995	0.338408	0.321441	0.109175	0.337164	0.214824	0.289061	0.306523	0.43241	0.469229	0.300223
37	35	0.184025	0.338448	0.321499	0.109204	0.337224	0.214862	0.289167	0.306593	0.432552	0.469381	0.300295
38	36	0.18405	0.338482	0.321548	0.109228	0.337276	0.214893	0.289256	0.306653	0.432673	0.46951	0.300357
39	37	0.184072	0.338511	0.321591	0.109249	0.337319	0.21492	0.289332	0.306704	0.432775	0.46962	0.300409
40	38	0.18409	0.338535	0.321626	0.109267	0.337356	0.214943	0.289397	0.306747	0.432862	0.469713	0.300454
41	39	0.184106	0.338556	0.321657	0.109282	0.337388	0.214962	0.289451	0.306784	0.432936	0.469793	0.300491
42	40	0.184119	0.338574	0.321683	0.109295	0.337414	0.214978	0.289498	0.306815	0.432999	0.46986	0.300524
43	41	0.184131	0.338589	0.321705	0.109306	0.337437	0.214992	0.289537	0.306842	0.433052	0.469918	0.300551
44	42	0.18414	0.338601	0.321723	0.109315	0.337457	0.215004	0.289571	0.306864	0.433098	0.469967	0.300574
45	43	0.184148	0.338612	0.321739	0.109323	0.337473	0.215014	0.289599	0.306883	0.433136	0.470008	0.300594
46	44	0.184155	0.338621	0.321753	0.109329	0.337487	0.215023	0.289623	0.306899	0.433169	0.470043	0.30061
47	45	0.184161	0.338629	0.321764	0.109335	0.337499	0.21503	0.289644	0.306913	0.433197	0.470073	0.300625
48	46	0.184166	0.338636	0.321774	0.10934	0.337509	0.215036	0.289661	0.306925	0.43322	0.470099	0.300637

	K	L	M	N	O	P	Q	R	S	T	U
		1	2	3	4	5	6	7	8	9	10
1	1	1	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	0	1	0	0	0	0
3	0	0	1	0	0	1	0	1	0	1	
4	1	0	0	1	0	0	0	0	0	0	0
5	0	0	0	0	1	1	0	0	1	1	
6	0	1	0	1	0	1	0	0	0	0	0
7	1	0	0	0	0	0	1	1	0	0	0
8	1	1	0	0	0	0	0	1	0	0	0
9	1	0	1	1	0	0	1	0	1	0	0
10	0	0	0	0	1	1	0	0	1	1	

Clearly from this figure it is apparent that some individuals (9 & 10) are far more likely to become infected than others (4 & 1). In fact, you can see in the figure below that the distribution of infection likelihood is dramatically different for almost every member of this population. The large ballooning of the infection rates for various cells shows exactly how much the underlying connections can affect individual infection rates even with the same starting infection rate.

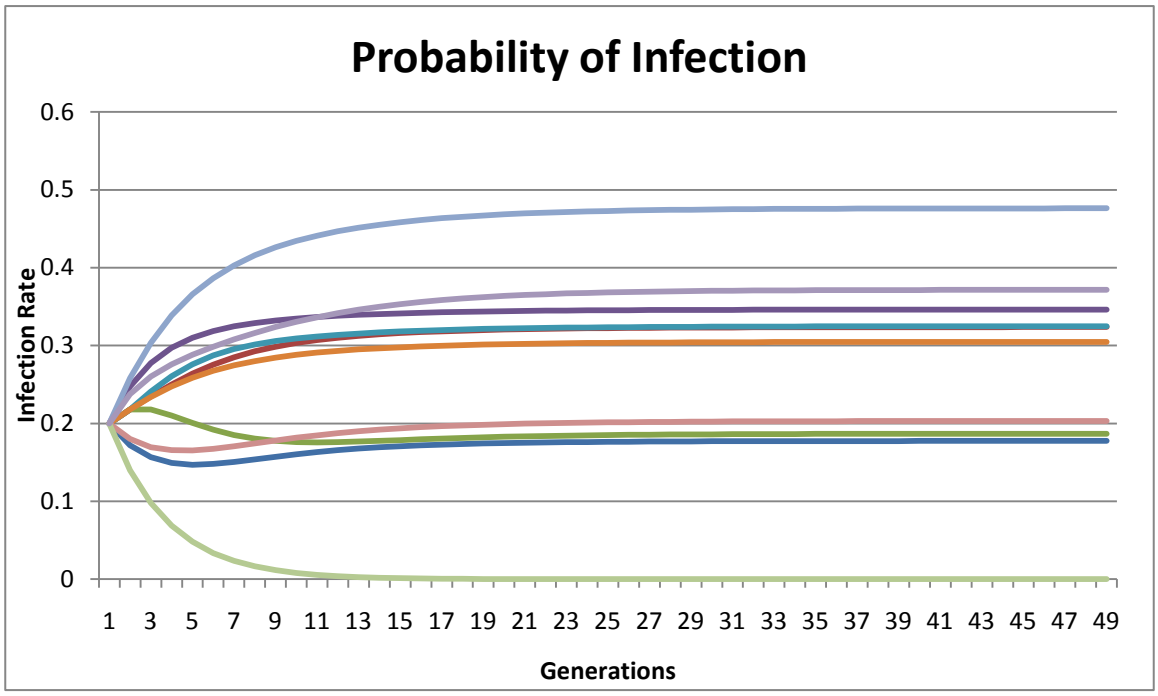


Figure 15 Individual Infection Rates

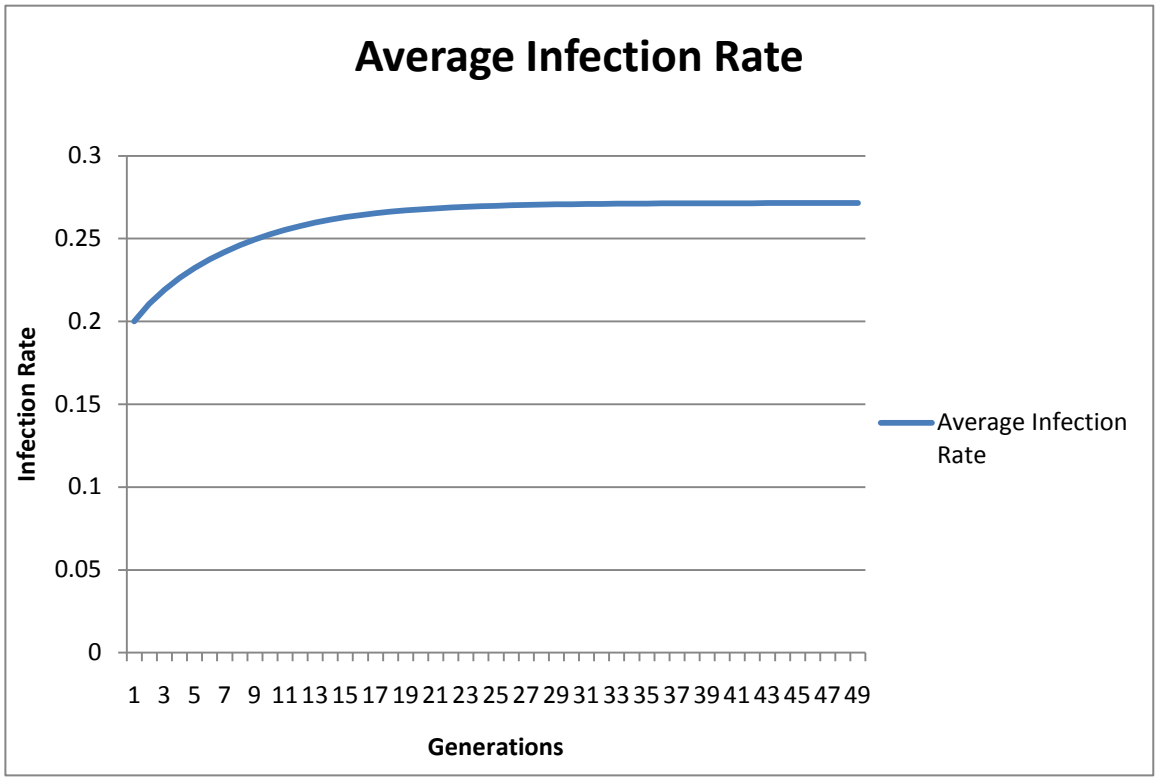


Figure 16 Average Infection Rates

The graph of the overall average infection rate shows none of the variability that Figure 16 above shows. While being able to compute average infection rates for a population may be useful for some statistical inferences, it really serves only to mask the large amount of variability that exists in the population. This variability is important as we have seen and can result in second and third order effects that would not be noticed with vast amounts of averaging.

V. CONCLUSION AND RECOMMENDATIONS

From the information flow scenario and disease transmission scenario we can glean two important lessons. The first is that the distribution of the connections in a network can have a substantial impact on the diffusion of something, say information, through a network. The second is that the connectedness of a network may also have a large effect on the spread of something, say a disease, through a network. These two facts highlight the importance of the underlying graph that represents a network when modeling movement of an impulse through that network. This impulse may be a message or it may be a virus but it is clear from these experiments that random mixing is not always the best assumption.

There are certainly times when random mixing is a good assumption to make. Within a house for example it could reasonably be expected that individuals would moving about the house fairly often and interacting with all individuals in the house at some point or another. This would not be true on a national scale however because you would not expect a person on one side of the country to interact often with individuals on another side of the country. In such a situation it would be best modeled using a simulation that took into account something about the underlying distribution of connections between people. This paper should serve to illustrate the idea that random mixing models may not work well for many situations. Additionally this paper presents an alternate way to simulate spread throughout a network if something about the underlying network graph is known.

A. Recommendations for further study

There are many useful tools that can be derived from further research into this problem of modeling diffusion. First, in order to refine the current modeling using CA, it would be useful to model real life networks or at least random networks that are similar to real ones. There is currently a lot of research in the area of pseudo-random graph generation and this project would be more complete if it incorporated some of the findings of that research. The random networks that were generated in this project are completely random and do not properly simulate the social networks one might find in a community of people. For example, random small-world networks would better model a social network but that would also require substantially more coding and was not able to be incorporated into this project.

A second recommendation is further research into the development of theoretical models of diffusions. This paper offers a simulation tool which can model diffusion but it does not lend itself nicely to strictly theoretical mathematical models. A single formula or series of equations to determine long term infection rates would be considerably more useful than a computer simulation. Additional research into this project area would first focus on obtaining a theoretical model for diffusion.

Another interesting idea to study in the future is the pinpointing of changes in behavior based on topology. It is clear in this paper that certain changes in topology do have an effect on diffusion. Finding out exactly which changes cause the change is another area yet to be researched. It would be interesting and useful to know exactly where the tipping points are that cause a diffusion model to exhibit such different results. Knowing the tipping points would indicate what kind of diffusion behavior could be expected from the onset when examining a particular network configuration.

REFERENCES

Alberts, David S, John J Garstka, and Frederick P Stein. "Network Centric Warfare: Developing and Leveraging Information Superiority." *Command and Control Research Program.* 1999. http://www.dodccrp.org/files/Alberts_NCW.pdf (accessed October 19, 2009).

Chestnut, Galanthay. *Stochastic Epidemic Modeling.* 2010. http://mathbio.colorado.edu/index.php/APPM4390:Stochastic_Epidemic_Modeling (accessed March 28, 2010).

Ediger, Bruce. *Simulating Network Worms.* 2010. http://www.stratigery.com/nws/ (accessed March 27, 2010)

Jackson, Matthew O. *Social and Economic Networks.* Princeton: Princeton University Press, 2008.

Kempe, David, Jon Kleinberg, and Eva Tardos. "Maximizing the Spread of Influence through a Social Network." *AMC Portal.* 2003. http://delivery.acm.org/10.1145/960000/956769/p137-kempe.pdf?key1=956769&key2=3458616521&coll=GUIDE&dl=GUIDE&CFID=57661957&CFTOKEN=12775274 (accessed October 21, 2009).

Weisstein, Eric W. *Kermack-McKendrick Model.* http://mathworld.wolfram.com/Kermack-McKendrickModel.html (accessed October 28, 2009).